



"THEY DO NOT KNOW THAT THEY DON'T KNOW": REVEALING AND QUANTIFYING THE SOCRATES BIAS

*Vasileios Kiosses¹ | Claire de Burbure² | Ioannis D K Dimoliatis³

¹ PgDip, PhD(c) in Medical Education, Medical Education Unit, Department of Hygiene and Epidemiology, Medical Faculty, School of Health Sciences, University of Ioannina, Ioannina, Greece, PC. 45110. (*Corresponding Author)

² Medical Education Advisor, Faculty of Medicine, Université catholique de Louvain, Brussels, Belgium.

³ Associate Professor of Public Health and Medical Education, Medical Education Unit, Department of Hygiene and Epidemiology, Medical Faculty, School of Health Sciences, University of Ioannina, Ioannina, Greece.

ABSTRACT

The objective of the present study was to determine whether bias (over- or under-estimation of self-competence) affects pre-training ratings and hence distorts the actual participation effect of experiential workshops.

Assessments were held during "empathy in doctor-patient relationship" elective courses held during winter 2014, spring 2015 and winter 2016 at Ioannina's Medical School, University of Ioannina, Greece.

Twenty-eight women and 19 men aged 21-28 years (mean = 22.8, SD = 1.52), in 4th (n = 18), 5th (n = 19) and 6th (n=10) year of medical studies took part, voluntarily, in the empathy training.

The Jefferson Scale of Physician Empathy was used on a total of 47 medical undergraduates to measure empathic performance both before (B) and a-posteriori-before (P) training. Overestimation of empathic ability was calculated as the difference B-P, and its significance was checked through paired t-test, while effect size (Cohen's d) was used to reveal any practical importance.

Participants' mean B score (+SD) was 110.6 (10.5) whereas P was 88.6 (13.8; p(B-P) < 0.001). Assuming total P as the basis (100), total B was 124.8, i.e 24.8% overestimation. A very large effect size was found (d = 1.81) for B-P indicating a highly practical importance. There were no significant differences between the 3 cohorts nor between men & women.

This study revealed the existence of the "do not know that they don't know" bias, offered a simple and easy method to measure it, and estimated it to be 24.8%.

KEYWORDS: the Socrates bias, medical education, self-assessment, questionnaire, bias.

INTRODUCTION

All self-assessment questionnaires are answered through the perceived reality of the participants in any search of their perceptions. The aspect that we perceive reality in a distorted way is not a revolutionary one. The philosophical dimension of the attitude that people are not aware of their ignorance, because they believe what they perceive through their senses constitutes the objective reality (the truth), was first said by Socrates in Plato's work "*Republic*" (514a-517c) (Bloom, 1968). Plato, influenced by the Socratic teaching, realized that in fact it is incorrect for every human to believe that he/she objectively knows the world. Everyone, through personal assessments and perceptions, knows subjectively the world. According to the Socratic and hence the Platonic philosophy, senses inevitably influence every judgment. Plato's "allegory of the cave" describes perfectly the imaginary world in relation to the sensible world.

The way in which we perceive reality may constitute a bias in research and affect the results and, if this kind of bias is not controlled or calculated, the results may be inaccurate. Cognitive bias, that may lead to distorted judgment, errors in decision making and illogical interpretation (Gilovich and Griffin, 2002), refers to "*a pattern of deviation in judgment, whereby inferences about other people and situations may be drawn in an illogical fashion*" (Hazelton et al., 2005).

At the Medical Education Unit, Medical School, University of Ioannina, Ioannina, Greece, we designed an empathy training course for medical undergraduates, the "*Empathize with me, Doctor!*" (*EwMD!*) project, aiming to improve students' empathic performance during their encounters with patients. The effectiveness of the training was assessed through the self-reported Jefferson Scale of Physician Empathy (JSPE). Before and after measurements revealed statistically significant improvements in empathic performance, which remained intact for at least six months (Kiosses et al., 2017). During the first group training, and hence students' ratings, we were very impressed with their before score (B), being almost 80% of the maximum JSPE score. Initially we thought that this was a JSPE inability to correctly measure empathic performance. After a while, a second idea came to our mind: perhaps they overestimated their empathic ability before training due to their ignorance of it, as students had indeed never been taught empathy so far. We called it the "they do not know that they don't know hypothesis" (DNKDNK). We then developed a method to estimate its magnitude and applied it to the following training groups. Results are presented here.

MATERIALS AND METHODS

Medical undergraduate volunteers from fourth to sixth year of studies at the University of Ioannina, Ioannina, Greece, successfully completed the experiential *EwMD!* training in three different small groups (winter semester 2014, spring semester 2015 and winter semester 2016). No ethical approval was needed because the training was conducted during the elective course "Empathy during doctor-patient relationship" at the Medical School, University of Ioannina, Greece. Details about the sample are described elsewhere (Kiosses et al., 2017).

Three training weekends, lasting 20 hours each, four weeks apart from each other, constituted the 60-hour training program, based on the principles of the Person-Centred Approach (PCA). The program was approved by the General Assembly of the Ioannina University Medical School (756/26-5-2013). The training included the principles and the implications of empathy during encounters with patients. The content of the training was wide including theory of the PCA, experiential active listening exercises, medical history taking, breaking bad news, exercises through art and play, use of open and closed questions and much more. A specific and detailed presentation of the content of the training is described elsewhere (Kiosses et al., 2017).

The 60-hour training program lasted two months and all medical undergraduates at the beginning of the training signed a declaration of consent. None of the undergraduates had ever participated in a similar training.

The Jefferson Scale of Physician Empathy (JSPE), a self-report inventory of 20 items rated on a 7-point Likert scale (strongly disagree to strongly agree) with higher scores indicating better empathic behaviour, was used to assess participants' empathic performance. JSPE is validated within the Greek population (Ouzouni and Nakakis, 2012) and its score ranges from 20 (worst) to 140 (best).

All trainees completed the JSPE anonymously. In order to match their ratings, they were asked to use a code, known exclusively to themselves. On the last day of the training, immediately after training completion, participants were asked to recomplete the JSPE using the same code, with the following instruction: "*With keeping in mind your current a-posteriori knowledge on what empathy actually is and how it can be gained, please complete again the questionnaire as if it was the day before your first day of the training*". This a-posteriori-before rating was

then compared with the before rating, using paired t-test. One way ANOVA was used to calculate any difference between cohorts. Chi square was used to identify any differences between men and women's ratings. The effect size was also calculated to assess the practical importance of the findings, using Cohen's d , interpreted as a small effect size if $d < 0.2$, medium if $0.2 \leq d \leq 0.5$, large if $d > 0.5$ (Cohen, 1988). The sum of before minus a-posteriori-before differences was used as an estimator of the magnitude of the DNKDNK bias. SPSS v.18 software was used (SPSS Inc.).

RESULTS

Forty-seven medical undergraduates took part in the study, 28 women and 19 men, aged 21 to 28 years (mean 22.8, SD 1.52), from 4th ($n = 18$), 5th ($n = 19$) and 6th ($n = 10$) year of study (see Table 1 for individual details). As shown in Table 2, participants' mean score (and standard deviation) before training (B) was 110.6 (10.5), whereas the a-posteriori-before (P) score was 88.6 (SD 13.8), indicating a highly significant difference (paired t-test with 46 degrees of freedom 11.35, $p < .001$, Table 2). This difference was observed to be equally highly significant in all studied subgroups, whether students are arranged by gender, by age or by semester of study (see Table 2). Interestingly, there were no statistically significant differences between men and women overall ($\chi^2 = 1.38$, $p = 0.24$, not shown). One way ANOVA revealed no significant difference between age groups and various cohorts for either before or a-posteriori before measurements ($F = 0.042$, $p = 0.96$ and $F = 0.557$, $p = 0.58$ respectively, not illustrated). At individual level, all trainees' P score was lower than their B score. A very large effect size was also observed for the B versus P measurement ($d = 1.81$), indicating a highly practical importance. The sum of all (B-P) differences was 1033 (see Table 2), a 24.8% increase if considering the sum of all P scores (4165) as the basis (100), hence the sum of all B scores (5198) became 124.8%. Figure 1 depicts the difference between the before and the a-posteriori-before rating. The between the two curves area (24.8% if the under the a-posteriori-before curve area is considered as the basis 100%), represents the "They do not know that they don't know" bias (DNKDNK).

DISCUSSION

We revealed the existence of the "They do not know that they don't know" (DNKDNK) bias, offered a simple and easy method to measure it, and estimated it to be 24.8% in a student-selected course at Ioannina University Medical School. Forty-seven medical undergraduates participated in a 60-hour experiential empathy training aiming at improving their empathic understanding during their encounters with their patients (the EwMD! project). When undergraduates were asked to reassess themselves after the last day of the training "as if it was the day before the first day of the training, keeping in mind their after the training knowledge on what empathy actually is and how it can be promoted" (a-posteriori-before self-assessment), they scored very much lower ($p < 0.001$) than their before rating (self-assessment), indicating that they, before training, had overestimated (by 24.8%) their abilities in being empathic, and they falsely believed that they knew how to promote an empathic condition during their clinical practice. Our method thus revealed the existence of the DNKDNK bias and offered a method to quantify it. In other words, trainees, during their participation in the EwMD! Project, had the chance to become aware of their own unawareness. After their participation, they were more sensitive about the philosophy of empathy, they were more aware about its use and its importance, but mostly they learned how to define it. Hence, without any external examiners, observers or specialists to rate them, they filled the JSPE on their own, rating mostly their unawareness.

Why and how did this occur? According to the philosophical stream of phenomenology, there is not one absolute objective reality which all humans perceive in the same way. The subjective way in which each person perceives and experiences the reality determines the behaviour of each human (Rogers, 1951). The question is why someone has such a distorted sense of his or her knowledge? Most medical undergraduates who took part in the training during this study said that they were disappointed when they realized that they were not as empathic as they thought they were. An explanation for this is that, according to the Person-Centred Approach, when a person has experiences that threaten his or her self-image, pertaining to all the characteristics used to describe oneself, then that person denies or distorts them. For example, it may be threatening for a person to realize that he/she is not empathic or cannot relate effectively with others. In order to fit this experience to the self-image, a person will distort or deny this perception and consequently he/she will be unaware of this incompetence.

It is also interesting to note the fact that there was no significant difference observed between men and women. This is consistent with previous studies observing no gender differences when participants were asked to complete a task equally relevant to men and women, then estimate their performance (Kim et al., 2015).

Additionally, as shown in Table 2, the Socrates bias exists in every participant regardless of their age or year of study. This unawareness of incompetence occurred in every cohort comparison within our sample. However, there was one subgroup for whom a marginally significant p value was observed (winter 2016, $p = 0.044$). This significant difference might be the result of chance alone. Indeed, if several independent null hypotheses were tested, the chances of obtaining at least one "statistically significant" result are greater than 5% (even if

all null hypotheses were true). This is the multiple comparison effect (Miller, 1981).

As we can see in Figure 1, one participant rated himself very strictly during the a-posteriori measurement. A possible interpretation of this finding may be that this student rated himself very strictly, after realizing what empathy actually was, and how it could be accurately promoted. After the completion of the EwMD!, the student realized the extent of his ignorance, yet overestimated his abilities and therefore tended to annihilate his adequacy previous to participating in the training.

Findings of the present study are consistent with other findings indicating that unskilled people are unaware of their incompetence. Specifically Ehrlinger et al. (2008) have found that poor performers lack insight into their incompetence and tend to be optimistic when evaluating their performance or knowledge. Furthermore, another study examined how physicians self-assess their competences compared with external observations. This systematic review revealed a minor relationship between self-assessment and external assessment. Hence the study suggested that physicians lack the ability to accurately self-assess (Davis et al., 2006).

Another interesting study compared results from three meta-analyses about medical students' self-assessment of performance. Findings indicated that medical students were more accurate in self-assessment and tended not to over- or underestimate their skills. However, medical students were less accurate in self-assessing their communication-based abilities, such as empathy, or abilities on encountering patients, tending to overestimate their competences. Knowledge-based performance self-assessments thus tended overall to be more accurate (Blanch-Hartigan, 2011).

According to Schiekirka et al. (2014) we need to take into consideration the impact of using retrospective measurements in medical education. Their research revealed that these retrospective ratings were more pessimistic than true pretest ratings, indicating some impact of response shift on students' self-assessments, although this impact was small.

The Dunning-Kruger effect is also a social bias where unskilled individuals suffer from illusory superiority and where highly competent people distort the abilities of others (Kruger & Dunning, 1999). The erroneous perception of the others creates a sense of incompetence in themselves. The difference between the Dunning-Kruger effect and the Socrates bias is that the Socrates bias refers to the sense of oneself, not according to others but according to the perceived reality in relation to the self. Trainees were self-assessed, and it was therefore not an external examiner, i.e. a presupposed more objective judge, who rated them, unlike the Dunning-Kruger effect. The Socrates bias does not accept the one and only truth, but the truth revealed in each human, by his/her own experience, and not in comparison to others. Additionally, the Dunning-Kruger effect uses external specialists in order to measure the differences, accepting that there is a subjective reality, while the Socrates bias measurement is based on each individual's perception of ignorance.

This study furthers knowledge not only in that bias occurs in medical education, but it furthermore offers an effective way to measure it, with no external examiner or specialist, but with the trainees themselves. In honour of the great philosopher Socrates, this bias is called "the Socrates bias" meaning the unawareness of ignorance.

Limitations

Measurement of retrospectiveness might be a limitation (recall bias) but we still believe that the Socrates bias would still occur even if the recall bias was controlled.

Future Research

Future studies could potentially use the Socrates bias in order to measure if this bias occurs in their research.

REFERENCES

1. Bloom, A. (1968). The Republic, Basic Books, New York.
2. Blanch-Hartigan, D. (2011). Medical Students' Self-Assessment of Performance: Results from Three Meta-Analyses. Patient Education and Counseling, 84(1), p.3-9.
3. Cohen, J. (1988). Statistical Power Analysis for the Behavioral Sciences, 2nd ed. Lawrence Erlbaum Associates, Hillsdale NJ.
4. Davis, A. D., Mazmanian, P. E., Fordis, M., Van Harrison, R., Thorpe, K. E., and Perrier, L. (2006). Accuracy of Physician Self-assessment Compared With Observed Measures of Competence: A Systematic Review. Journal of American Medical Association, 299(9), p. 1094-1102.
5. Ehrlinger, J., Johnson, K., Banner, M., Dunning, D. and Kruger, J. (2008). Why the unskilled are unaware: Further explorations of (absent) self-insight among the incompetent. Organizational Behavior and Human Decision Processes, 105(1), p.98-121.
6. Gilovich, T. and Griffin, D. W. (2002). Heuristics And Biases: Then And Now, in Heuristics and Biases: The Psychology of Intuitive Judgment, Gilovich, T., Kahneman, D. (ed), Cambridge University Press. Cambridge, p. 695-698.
7. Hazelton, M. G., Nettle, D., and Andrews, P. W. (2005). The Evolution Of Cognitive

- Bias, in: The Handbook of Evolutionary Psychology, Buss, D. M. (ed), John Wiley & Sons Inc. Hoboken NJ, p. 968-987.
8. Kim, Y. H., Chiu, C. Y., and Bregant, J. (2015). Unskilled and Don't Want to Be Aware of It: The Effect of Self-Relevance on the Unskilled and Unaware Phenomenon. *PLoS ONE*, 10(6).
 9. Kiosses, V., Tatsioni, A., Dimoliatis, I., and Hyphantis, T. (2017). "Empathize With Me, Doctor!" Medical Undergraduates Training Project: Development, Application, Six-Month Follow-Up. *Journal of Education and Training Studies*, 5(7), p.20-27.
 10. Kruger, J., and Dunning, D. (1999). Unskilled and Unaware of It: How Difficulties in Recognizing One's Own Incompetence Lead to Inflated Self-Assessments. *Journal of Personality and Social Psychology*, 77(6), p.1121-34.
 11. Miller, R.G. (1981). *Simultaneous Statistical Inference* 2nd Ed. Springer Verlag, New York.
 12. Ouzouni, C., and Nakakis, K. (2012). An Exploratory Study Of Student Nurses' Empathy. *Health Science Journal*, 6(3), p.534-552.
 13. Rogers, C. R. (1951). *Client-Centered Therapy: Its Current Practice, Implications, And Theory*, Houghton Mifflin Co, Boston.
 14. Schierkirk, S., Anders, S., and Raupach, T. (2014). Assessment Of Two Different Types Of Bias Affecting The Results Of Outcome-Based Evaluation In Undergraduate Medical Education. *BMC Medical Education*, 14, 149.

Table 1.
Jefferson Scale of Physician Empathy (JSPE) scores before (B) and a-posteriori-before (P) by rising Socrates bias (B-P) per Student (IDentification, Gender [Male, Female], Age, Year of study, Semester Course [w14 winter 2014, s15 spring 2015, w16 winter 2016]).

Nr	ID	Gender	Age	Year	Sem	Before (B)	a-Posteriori-before (P)	Socrates Bias (B-P)
1	K	F	28	6	w14	107	102	5
2	F	M	25	6	s15	103	97	6
3	D	F	24	6	s15	101	93	8
4	G	F	23	5	s15	104	96	8
5	H	F	21	4	w14	106	98	8
6	J	M	23	4	w14	107	99	8
7	P	M	22	5	w14	115	107	8
8	B	M	23	5	s15	91	82	9
9	C	M	23	5	s15	101	92	9
10	Q'	M	21	4	w16	98	88	10
11	F'	F	22	4	w16	81	70	11
12	J'	F	21	5	w16	93	82	11
13	N'	F	22	5	w16	113	101	12
14	N	F	22	4	s15	114	101	13
15	P'	F	23	5	w16	105	92	13
16	T'	M	22	4	w16	115	100	15
17	Z	F	22	5	w14	124	108	16
18	H'	F	21	4	w16	103	86	17
19	W	M	24	6	w14	120	102	18
20	O	F	23	5	s15	115	95	20
21	U	M	24	5	w14	119	99	20
22	V	F	23	6	w14	119	99	20
23	R	M	23	5	s15	117	96	21
24	A	F	25	6	s15	87	65	22
25	U'	F	22	5	w16	104	81	23
26	C'	F	23	6	w14	128	104	24
27	E	F	22	4	s15	102	78	24
28	M	F	22	4	w14	114	90	24
29	Q	F	22	4	s15	117	93	24
30	S	F	21	4	w14	118	94	24
31	I'	F	21	4	w16	116	91	25
32	L'	M	24	5	w16	101	76	25
33	R'	M	22	4	w16	107	81	26
34	K'	F	23	5	w16	112	85	27
35	X	F	22	4	s15	122	95	27
36	L	F	23	5	s15	109	80	29
37	O'	F	22	5	w16	113	84	29
38	S'	M	21	4	w16	112	83	29
39	M'	F	22	5	w16	114	84	30
40	B'	M	23	5	w14	125	94	31
41	G'	M	22	4	w16	109	78	31
42	A'	M	22	4	s15	125	93	32
43	T	F	26	6	s15	119	87	32
44	Y	M	22	4	w14	123	87	36
45	D'	F	27	6	s15	129	91	38
46	E'	M	22	5	w16	115	57	58
47	I	M	24	6	w14	106	29	77

Table 2.

Group analysis. Sum of JSPE scores given by participants (n); in parentheses, percentage of sum, choosing P sum as basis (100). Mean score (standard deviation). Socrates bias (B-P), and its p-values. Comparisons of two subgroups for each group (do subgroups differ? raw p-values).

	Before (B)	a-Posteriori- before (P)	Socrates bias (B-P)	p-value †
Total number of Students				
n	47	47	47	
Sum (%)	5198 (124.8)	4165 (100)	1033 (24.8)	
Mean (SD)	110.6 (10.5)	88.6 (13.8)	22.0 (13.3)	< 0.001 *
Female Students				
n	28	28	28	
Sum (%)	3089 (122.3)	2525 (100)	564 (22.3)	
Mean (SD)	110.3 (11.2)	90.2 (10.0)	20.1 (8.5)	< 0.001
Male Students				
n	19	19	19	
Sum (%)	2109 (128.6)	1640 (100)	469 (28.6)	
Mean (SD)	111.0 (9.6)	86.3 (18.1)	24.7 (18.1)	< 0.001
p-value ††	0.830578	0.352756	0.318825	
21-22 year-old Students				
n	25	25	25	
Sum (%)	2773 (125.5)	2210 (100)	563 (25.5)	
Mean (SD)	110.9 (10.1)	88.4 (11.5)	22.5 (11.1)	< 0.001
23-28 year-old Students				
n	22	22	22	
Sum (%)	2425 (124)	1955 (100)	470 (24)	
Mean (SD)	110.2 (11.1)	88.9 (16.4)	21.4 (15.6)	< 0.001
p-value ††	0.824242	0.910137	0.769267	
4th Year Medical Students				
n	18	18	18	
Sum (%)	1989 (123.9)	1605 (100)	384 (23.9)	
Mean (SD)	110.5 (10.5)	89.2 (8.6)	21.3 (8.7)	< 0.001
5th + 6th Year Medical Students				
n	29	29	29	
Sum (%)	3209 (125.4)	2560 (100)	649 (25.4)	
Mean (SD)	110.7 (10.7)	88.3 (16.4)	22.4 (15.6)	< 0.001
p-value ††	0.589955	0.944749	0.618142	
Winter 2014 + Spring 2015 Cohorts				
n	30	30	30	
Sum (%)	3387 (123.3)	2746 (100)	641 (23.3)	
Mean (SD)	112.9 (10.6)	91.5 (14.8)	21.4 (14.2)	< 0.001
Winter 2016 Cohort				
n	17	17	17	
Sum (%)	1811 (127.6)	1419 (100)	392 (27.6)	
Mean (SD)	106.5 (9.3)	83.5 (10.4)	23.1 (11.8)	< 0.001
p-value ††	0.044309	0.053651	0.679210	

† Based on the two tailed, paired t test, with n-1 degrees of freedom

†† Based on the two-tailed, two-sample equal variance (unequal variance if the largest is more than two times the smaller) t-test, with n1+n2-2 degrees of freedom

* Effect size estimation: Cohen's d = 1.81 (d < 0.2 small; 0.2 ≤ d ≤ 0.5 medium; d > 0.5 large)

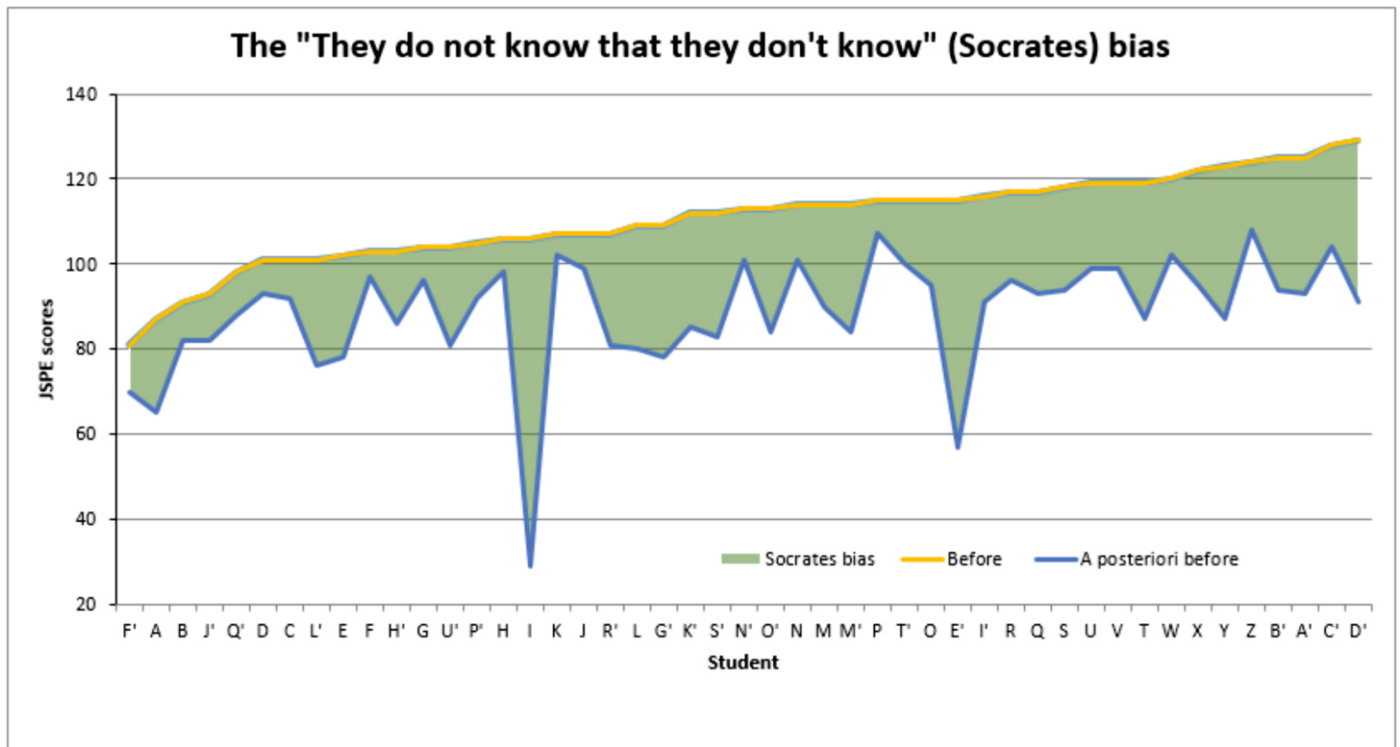


Figure 1. JSPE scores in the before (B) and a-posteriori-before (P) measurements for each student, sorted by before scores